

Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model

Rolf H. Reichle¹ and Randal D. Koster

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Received 7 October 2004; revised 29 November 2004; accepted 16 December 2004; published 22 January 2005.

[1] Global retrievals of surface soil moisture from the Scanning Multichannel Microwave Radiometer for the period 1979–87 are assimilated into the NASA Catchment land surface model as it is driven with surface meteorological data derived from observations. Validation against ground-based measurements in Eurasia and North America from the Global Soil Moisture Data Bank demonstrates a long assumed (but rarely proven) property of soil moisture fields derived from data assimilation – that the assimilation product is superior to either satellite data or model data alone. An analysis of the innovations reveals that the filter is only partially operating within its underlying assumptions and offers clues how spatially distributed model error parameters could further enhance filter performance. **Citation:** Reichle, R. H., and R. D. Koster (2005), Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, *Geophys. Res. Lett.*, 32, L02404, doi:10.1029/2004GL021700.

1. Motivation

[2] Global fields of the vertical profile of soil moisture are needed, for example, to initialize sub-seasonal forecasts of summer precipitation and air temperature over mid-latitude land [Koster *et al.*, 2004]. Yet in situ measurements of root zone soil moisture are limited to parts of Eurasia and North America. Alternatively, soil moisture can be derived from a land surface model forced with observed precipitation, radiation, and other surface meteorological data [Rodell *et al.*, 2003]. Additional information may be provided by satellite observations of C-band (6.6 GHz) or L-band (1.4 GHz) radiobrightness temperature, which can be interpreted in terms of surface soil moisture in the top 1 cm or 5 cm soil layer, respectively. It has long been argued – but not yet proven decisively – that assimilation of satellite retrievals of surface soil moisture into a land model provides superior estimates of global soil moisture conditions, even in the root zone (typically 1 m deep).

[3] While there has been considerable progress in the methodological development of soil moisture data assimilation [Walker and Houser, 2001; Margulis *et al.*, 2002; Reichle *et al.*, 2002; Crow and Wood, 2003; Seuffert *et al.*, 2003], there is little experience with the assimilation of a multi-year, global data set of surface soil moisture retrievals. In this paper, we assimilate global soil moisture retrievals

from the Scanning Multichannel Microwave Radiometer (SMMR) into the NASA Catchment land surface model [Koster *et al.*, 2000] for the period 1979–87. Through validation against ground measurements, we demonstrate that assimilation of SMMR data yields improved soil moisture estimates – better than those obtained with the model or from the satellite alone.

2. Data and Method

[4] Owe *et al.* [2001] recently developed a novel retrieval algorithm for soil moisture from passive microwave measurements and produced a nine-year, global soil moisture data set from SMMR observations for the period 1978–87 [De Jeu, 2003]. For the period 1979–93, Berg *et al.* [2003] developed a high-quality, global data set of surface meteorological fields based on reanalysis data and corrected with observations as much as possible. This data set is used to force the NASA Catchment land surface model. Finally, ground-based soil moisture data for the SMMR time period are available for select locations in Eurasia and North America from the Global Soil Moisture Data Bank (GSMDB) [Robock *et al.*, 2000].

[5] These satellite, ground-based, and model soil moisture data are independent, and each has its own set of limitations [Reichle *et al.*, 2004]. State-of-the-art land surface models produce widely different soil moisture output even when integrated with identical meteorological forcing inputs [Entin *et al.*, 1999]. Errors in C-band surface soil moisture retrievals are generally high. Modest amounts of vegetation obscure the soil moisture signal, which is limited to the top 1 cm of the soil. Up to 10 retrievals per month are available from SMMR (see Figure 1 (top) for regions where retrievals are typically available.) Ground-based measurements – used for validation – are sparse and not necessarily representative of large-scale soil moisture. At this time, errors in global soil moisture observation and modeling are so large that there is no universally accepted climatology. Consequently, we scale the satellite observations to the model's climatology before assimilating the data into the model [Reichle and Koster, 2004]. For seasonal climate prediction, knowledge of soil moisture anomalies is, in any case, more important than knowledge of absolute soil moisture.

[6] In a data assimilation system, the model-generated soil moisture is corrected toward the observational estimate, with the degree of correction determined by the levels of error associated with each. The assimilation system used here is based on the Ensemble Kalman filter (EnKF), which is well suited to the nonlinear and intermittent character of land surface processes [Reichle *et al.*, 2002]. The key

¹Also at Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland, USA.

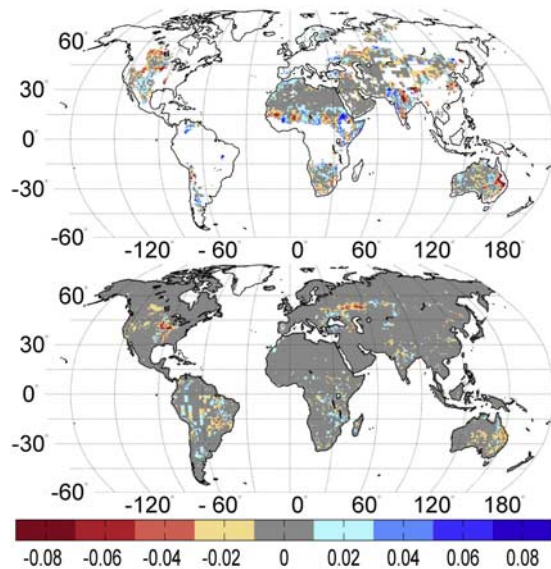


Figure 1. Difference fields for July 1983 anomalies of surface soil moisture [m^3m^{-3}]: (top) SMMR anomalies minus model anomalies, (bottom) EnKF anomalies minus model anomalies.

feature of the EnKF is that error estimates of the model-generated results are dynamically derived from an ensemble of model integrations. Each member of the ensemble experiences slightly perturbed instances of the observed precipitation fields (representing errors in the precipitation data) and is also subject to randomly generated noise that is directly added to the soil moisture states (representing errors in model physics and parameters). In this paper, we use the one-dimensional version of the EnKF. Preliminary results with the three-dimensional EnKF [Reichle and Koster, 2003] show a further improvement in surface soil moisture but also slightly less skill for root zone soil moisture. Calibration of the latter is very complex and work is still in progress.

[7] In the next section, we analyze “raw” time series of monthly mean soil moisture as well as anomaly time series. The latter are obtained by subtracting the monthly climatology of each data set (i.e., the average for each calendar month) from the raw time series. In other words, the raw time series include the seasonal cycle, while the anomaly time series describe only deviations from the average seasonal cycle. Our analysis focuses on time series correlations between the various data sets rather than on root-mean-square errors, because there is not enough evidence to tell whether the climatology of the ground-, satellite-, or model-based data is most correct [Reichle et al., 2004].

3. Results

[8] Figure 1 shows an example of the assimilation product. Figure 1 (top) shows the difference between the monthly mean anomaly fields of the SMMR retrievals and the model soil moisture for July 1983. For example, both the model and the (scaled) SMMR data suggest a wet anomaly in the northern Great Plains of North America during this month, but the anomaly is weaker in the SMMR data, and thus the difference in the anomalies shows up as negative in Figure 1. The assimilation algorithm is designed

to estimate the true anomalies by combining the model and SMMR data. In practice, the EnKF assimilation anomaly tends to lie between the model and SMMR estimates. This is evident in Figure 1 (bottom), which shows the EnKF anomaly minus the model anomaly. This difference is, to first order, a damped version of the differences shown in Figure 1 (top), as can be seen, for example, in North America, central Eurasia, and Australia. In the Sahel there is almost no response, suggesting that here the assimilation algorithm places much more weight on the model than on the SMMR observations.

[9] Figure 1 primarily demonstrates that the assimilation system works as designed. There are, however, some interesting dynamical effects worth noting. For example, the SMMR data also influence the EnKF estimates of root zone soil moisture (not shown), allowing the temporal propagation of SMMR information. In the midwestern USA (along -90 degrees longitude), SMMR data reduce the strength of the EnKF anomaly for root zone moisture early in the summer of 1983. The memory contained in this moisture reservoir then carries forward the weaker anomaly into July, when few SMMR data are available (most likely because vegetation grew too dense). As a result, Figure 1 shows that, despite the essential absence of SMMR data in the midwest during July 1983, the EnKF surface anomaly for that month differs from that of the model.

[10] The global performance of the assimilation algorithm is reflected, in part, in its innovations sequence (the difference between SMMR retrievals and their corresponding model forecasts during the assimilation integration). If the filter operates according to its underlying assumptions – that various linearizations hold, and that model and observation errors are uncorrelated and normally distributed – then the sum of the model error covariance (diagnosed from the ensemble spread) and the measurement error covariance should equal the sample covariance of the innovations sequence. In other words, we can check the assumptions underlying the assimilation process by checking whether the innovations sequence has the expected mean and variance [Reichle et al., 2002].

[11] Because of the bias reduction applied before the assimilation, the mean of the innovations is statistically indistinguishable from zero. A supplemental analysis shows that not scaling the SMMR data a priori leads to a mean that is about one standard deviation away from the expected mean of zero, providing further evidence that bias removal is an indispensable part of the assimilation system. The global average variance of the normalized innovations, shown in Figure 2, is around 0.7, somewhat short of the expected value of 1. Moreover, there are strong variations

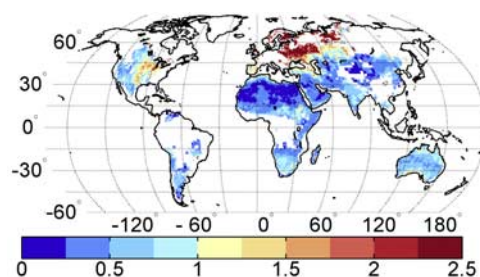


Figure 2. Variance of normalized innovations [-].

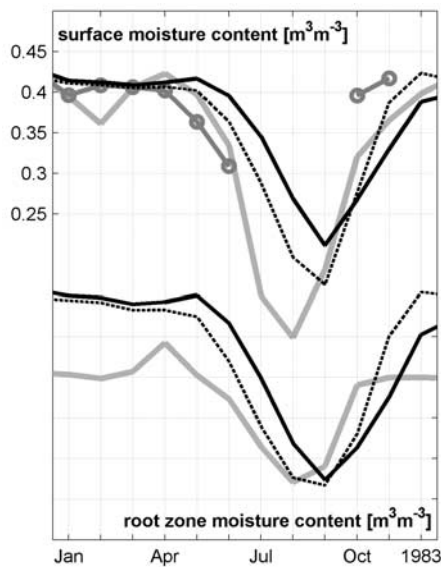


Figure 3. (Top group of lines, left axis) Surface and (Bottom group of lines, right axis) root zone soil moisture for 1983 at a representative location in Illinois (89.5W, 38.6N): (Light gray) GSMDB, (Dark grey with circles) bias-corrected SMMR, (Black solid) model, (Black dashed) EnKF.

across the globe. The variance slightly exceeds 1 in central North America, and it is closer to 2 in mid-latitude Eurasia. For the rest of the globe, the innovations show too small a variance. These imperfections are explained in part by nonlinearities in the model and in the observation operator. They also relate, however, to an imperfect representation of the model error characteristics in the ensemble generation. Since the variance of the normalized innovations is inversely related to the model forecast error variance, Figure 2 suggests too little (too much) model or rainfall error in wet (arid) climates. It might be possible to use the innovations variance to tune filter parameters (such as model error variances) before repeating the assimilation integration. Alternatively, adaptive tuning methods could be tried [Dee, 1995].

[12] Without any such tuning of the filter, we now show that the assimilation of SMMR retrievals already yields modest but significant improvements in the estimation of soil moisture. For this validation, we use in situ observations from up to 77 locations in North America and Eurasia that have sufficient GSMDB and SMMR data [Reichle *et al.*, 2004]. Figure 3 shows the monthly mean surface and root zone soil moisture for 1983 at a representative location

in Illinois. Note that the SMMR data (but not the ground data) have been corrected to the model climatology. At this location, the phase of the model data lags that of the ground data by about one month. The SMMR data, on the other hand, show a better phase agreement with the ground data, though the SMMR data are not available during summer because the vegetation then is too dense. The assimilation of just a few months of SMMR data per year shifts the spring dry-down and fall wet-up by about half a month towards the phase of the annual cycle of the ground data. Most importantly, this phase improvement applies equally to the root zone, where no SMMR data are available. Moreover, we see again (like in Figure 1) how memory in the root zone affects moisture in the surface layer. The monthly mean surface soil moisture of the assimilation integration stays below that of the model during July–September even in the absence of SMMR data.

[13] Table 1 provides a stronger, global-scale demonstration of improvement associated with assimilation. Listed are estimated time series correlation coefficients (R) with in situ data, computed from monthly mean time series and averaged over all locations with sufficient data in North America and Eurasia. Also given are 95% confidence intervals for R . (These must be interpreted with care for the raw time series because of the fixed-phase seasonal cycle signal.) For surface soil moisture, the satellite and model data show about the same skill in reproducing the in situ data, with R equal to 0.44 and 0.43, respectively (0.32 and 0.36, respectively, for anomalies.) Merging the SMMR retrievals with the model through data assimilation leads to an increase in R to 0.50 (0.43 for anomalies). This increase is highly statistically significant, with confidence levels of 99.7–99.9% based on a Monte-Carlo analysis (last two columns of Table 1). From the listed R 's and their uncertainties, we generated thousands of random pairs of R values, one of each pair representing R for the EnKF, and the other the R for SMMR or the model. We then computed the probability that the increase in R is real by determining the fraction of sample pairs with a positive difference in R .

[14] The model's skill for root zone soil moisture is comparable to its skill at the surface (Table 1), with R equal to 0.46 (0.32 for anomalies). Merging the surface information contained in the SMMR retrievals via data assimilation leads to a small increase in R for the root zone soil moisture to 0.50 (0.35 for anomalies). For root zone moisture, the increase is still statistically significant, albeit at a somewhat lower confidence level (97%). For the root zone anomalies, the confidence level for improvement due to the assimilation drops to 80%. This is partly because the increase in R

Table 1. Average Time Series Correlation Coefficients R With GSMDB Surface and Root Zone Soil Moisture (sfmc and rzmc, Respectively) for SMMR, Model, and Assimilation Estimates With 95% Confidence Intervals^a

	N	Correlation Coefficient With GSMDB Data [-]			Confidence Levels [%]: Improvement of EnKF	
		SMMR	Model	EnKF	Over SMMR	Over Model
Sfmc	77	.44 ± .03	.43 ± .03	.50 ± .03	99.7	99.9
Sfmc anomalies	66	.32 ± .03	.36 ± .03	.43 ± .03	99.9	99.9
Rzmc	59	-	.46 ± .03	.50 ± .03	-	97
Rzmc anomalies	33	-	.32 ± .05	.35 ± .05	-	80

^aN denotes the number of locations with sufficient data. Also shown are confidence levels that R for EnKF estimates is higher than R for SMMR (or model) data.

from 0.32 to 0.35 is smaller, and partly because there are fewer data available and thus the 95% confidence intervals around R are larger. Note also that even if the assimilation data were perfect, R could still be much less than 1 due to the mismatch of scale between the assimilation data and the GSMDDB data. In other words, the seemingly modest increase in R from 0.43 to 0.50 could be quite large relative to the maximum increase possible given the point-scale character of the validation data.

[15] The increase in time series correlations with in situ data after assimilation suggests that the satellite and model data contain independent information that the assimilation algorithm can combine into superior estimates. Note that the success of the assimilation system in improving estimates of root zone moisture, perhaps the key assimilation product, hinges on many factors. The model, for example, must accurately describe the propagation of the surface information into the deeper soil. Also, the model error parameters of the assimilation system that co-determine the strength of the coupling between the surface and the root zone must be realistic. Unfortunately, we cannot now test these assumptions at the global scale with any confidence. Furthermore, our ground validation data are point measurements and may be quite limited in their ability to represent soil moisture at the catchment scale. Still, despite these limitations, the assimilation of SMMR retrievals does yield improved estimates of soil moisture conditions.

4. Conclusions

[16] The global assimilation of SMMR satellite soil moisture retrievals into the NASA Catchment land surface model using the EnKF was examined. The assimilation improves the average annual cycle of surface and root zone soil moisture at locations with GSMDDB ground data. The assimilation also produces small but significant improvements in time series correlations with ground data for surface soil moisture and its anomalies. Correlations for root zone soil moisture are also improved, though not with the same statistical significance.

[17] Global analysis of the innovations sequence reveals that the assimilation algorithm only partially performs within its underlying assumptions. This is not surprising for such a first global assimilation of satellite data into a state (soil moisture) controlled by poorly understood non-linear processes. In future work, information from the innovations sequence will be used to design spatially distributed model error parameters, potentially in an adaptive framework, that might improve the performance of the assimilation algorithm. Finally, modern-era data such as C-band retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System, L-band retrievals from the planned Hydrospheric States mission [Entekhabi et al., 2004], and satellite-supported surface meteorological observations of higher quality should further our assimilation-based estimation of global soil moisture fields.

[18] **Acknowledgments.** We thank M. Owe and R. de Jeu for the SMMR data, A. Berg and J. Famiglietti for the forcing data, and W. Crow, S. Schubert, and one anonymous reviewer for helpful comments.

References

- Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser (2003), The impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes, *J. Geophys. Res.*, 108(D16), 4490, doi:10.1029/2002JD003334.
- Crow, W. T., and E. F. Wood (2003), The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97, *Adv. Water Resour.*, 26, 137–149.
- Dee, D. P. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, 123, 1128–1145.
- De Jeu, R. A. M. (2003), Retrieval of land surface parameters using passive microwave remote sensing, Ph.D. diss., Dep. of Geo-Environ. Sci., Vrije Univ., Amsterdam, Netherlands.
- Entekhabi, D., et al. (2004), The Hydrosphere State (HYDROS) mission concept: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *Trans. Geosci. Remote Sens.*, 42, 2184–2195.
- Entin, J. K., A. Robock, K. Y. Vinnikov, V. Zabelin, S. Liu, A. Namkai, and T. Adyasuren (1999), Evaluation of Global Soil Wetness Project soil moisture simulations, *J. Meteorol. Soc. Jpn.*, 77, 183–198.
- Koster, R. D., M. J. Suarez, A. Ducharme, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, 105, 24,809–24,822.
- Koster, R. D., M. J. Suarez, P. Liu, U. Jambor, A. A. Berg, M. Kistler, R. H. Reichle, M. Rodell, and J. Famiglietti (2004), Realistic initialization of land surface states: Impacts on subseasonal forecast skill, *J. Hydrometeorol.*, 5, 1049–1063.
- Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, *Water Resour. Res.*, 38(12), 1299, doi:10.1029/2001WR001114.
- Owe, M., R. de Jeu, and J. Walker (2001), A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, *IEEE Trans. Geosci. Remote Sens.*, 39, 1643–1654.
- Reichle, R. H., and R. D. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol.*, 4, 1229–1242.
- Reichle, R. H., and R. D. Koster (2004), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, 31, L19501, doi:10.1029/2004GL020938.
- Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, 3, 728–740.
- Reichle, R. H., R. D. Koster, J. Dong, and A. A. Berg (2004), Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation, *J. Hydrometeorol.*, 5, 430–442.
- Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai (2000), The global soil moisture data bank, *Bull. Am. Meteorol. Soc.*, 81, 1281–1299.
- Rodell, M., et al. (2003), The Global Land Data Assimilation System, *Bull. Am. Meteorol. Soc.*, 85, 381–394, doi:10.1175/BAMS-85-3-381.
- Seuffert, G., H. Wilker, P. Viterbo, J.-F. Mahfouf, M. Drusch, and J.-C. Calvet (2003), Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data, *Geophys. Res. Lett.*, 30(10), 1498, doi:10.1029/2003GL017128.
- Walker, J. P., and P. R. Houser (2001), A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *J. Geophys. Res.*, 106, 11,761–11,774.

R. D. Koster and R. H. Reichle, NASA-GSFC, Code 900.3, Greenbelt Rd, Greenbelt, MD 20771, USA. (reichle@gmao.gsfc.nasa.gov)